

CLAIMS

Claim 1. A method for generating a hash table for mapping sample reads to a reference sequence, the method comprising:

- obtaining, by a computer system, a first seed of nucleotides from a reference sequence, wherein the first seed has a length of K nucleotides;

- determining, by the computer system, that the first seed matches more than a predetermined number of reference sequence locations;

- based on determining that the first seed matches more than a predetermined number of reference sequence locations, generating, by the computer system, a seed extension tree having a plurality of nodes, wherein each node of the plurality of nodes corresponds to

 - an extended seed that is an extension of the first seed and has a nucleotide length of K^* , wherein K^* is one or more nucleotides greater than K, and

 - one or more locations, in a seed extension table, that include data describing reference sequence locations that match the extended seed; and

- for each node of the plurality of nodes:

 - storing, by the computer system, interval information at a location of the hash table that corresponds to an index key for the extended seed, wherein the interval information references one or more locations in the seed extension table that include data describing reference sequence locations that match the extended seed associated with the node.

Claim 2. The method of claim 1, wherein each of the matching reference sequence locations includes the K nucleotides of the first seed.

Claim 3. The method of claim 1, further comprising:

- obtaining, by the computer system, a second seed of nucleotides from the reference sequence that is different than the first seed;

- determining, by the computer system, that the second seed does not match more than the predetermined number of reference sequence locations; and

based on determining, by the computer system, that the second seed does not match more than the predetermined number of reference sequence locations

obtaining, by the computer system, data describing each of the reference sequence locations that match the second seed; and

storing, by the computer system, the data describing the reference sequence locations that match the second seed at a second location of the hash table that corresponds to an index key for the second seed.

Claim 4. The method of any one of claims 1 to 3, wherein the one or more locations in the seed extension table that include data describing reference sequence locations that match the extended seed comprises a contiguous interval of locations in the seed extension table that include data describing reference sequence locations that match the extended seed.

Claim 5. The method of any one of claims 1 to 3, wherein the one or more locations in the seed extension table that include data describing reference sequence locations that match the extended seed comprises a non-contiguous interval in an extension table of reference sequence locations that match the extended seed.

Claim 6. The method of any one of claims 1 to 5, wherein obtaining, by a computer system, a first seed of nucleotides from a reference sequence, the first seed representing a sequence of nucleotides having a nucleotide length of K nucleotides comprises:

determining, by the computer system, a location of a seed access window within a reference sequence; and

obtaining, by the computer system, a subset of the reference sequence that is identified by the seed access window.

Claim 7. The method of claim 6, further comprising:

adjusting, by the computer system, the seed extension window K nucleotides forward along the reference sequence to identify a second seed of nucleotides from the reference sequence having a nucleotide length of K nucleotides;

obtaining, by the computer system, the second seed from the reference sequence;

determining, by the computer system, that the second seed matches more than a predetermined number of reference sequence locations;

based on determining that the second seed matches more than a predetermined number of reference sequence locations, generating, by the computer system, a second seed extension tree having a plurality of second nodes, wherein each second node of the plurality of second nodes corresponds to (i) a second extended seed that is an extension of the second seed and has a nucleotide length of K^* , wherein K^* is one or more nucleotides greater than K , and (ii) a second one or more locations, in a second seed extension table, that includes data describing reference sequence locations that match the second extended seed; and

for each second node of the plurality of second nodes:

storing, by the computer system, second interval information at a location of the hash table that corresponds to an index key for the second extended seed, wherein the second interval information references one or more locations in the second seed extension table that include data describing reference sequence locations that match the second extended seed associated with the second node.

Claim 8. The method of any one of claims 1 to 7, further comprising:

for each node of the plurality of nodes:

determining, by the computer system, whether the node of the seed extension tree is a leaf node; and

based on determining, by the computer system, that the node of the extension tree is not a leaf node, storing, by the computer system, an extend record in the location of the hash table that corresponds to the index key for the extended seed.

Claim 9. The method of claim 8, wherein the extend record includes one or more instructions that, when executed by the computer system, cause the computer system to add one or more additional nucleotides to a seed associated with the extend record.

Claim 10. The method of any one of claims 1 to 7, further comprising:

based on determining, by the computer system, that the node extension tree is a leaf node, determining, by the computer system, not to store an extend record in the location of the hash table that corresponds to the index key for the extended seed.

Claim 11. The method of any one of claims 1 to 7, further comprising:
generating, by the computer system, the seed extension table, comprising:
identifying, by the computer system, each seed of the reference sequence that matches the first seed; and
storing, by the computer system, data identifying the identified seeds in the seed extension table.

Claim 12. The method of any one of claims 1 to 11, further comprising:
sorting, by the computer system, the identified seeds in the seed extension table.

Claim 13. The method of any one of claims 1 to 12, further comprising:
generating, by the computer system, a hash table installation package, the hash table installation package including instructions that, when processed by one or more computers that receive the hash table installation package, cause the one or more computers to install the hash table in a memory that is accessible by a programmable logic circuit.

Claim 14. The method of claim 13,
wherein the hash table installation package includes the seed extension table, and
wherein the hash table installation package includes instructions to instruct (i) the programmable logic circuit or (ii) another computer, to store the seed extension table in the memory device that is accessible to the programmable logic circuit.

Claim 15. The method of any one of claims 13 or 14, the method further comprising providing, by the computer system, the hash table installation package to another computer.

Claim 16. The method of claim 15, wherein the other computer includes (i) a computer that is configured to communicate with the programmable logic circuit or (ii) the programmable logic circuit.

Claim 17. The method of any one of claims of claim 1 to 16, wherein the computer system comprises a plurality of computers.

Claim 18. A system for generating a hash table for mapping sample reads to a reference sequence comprising:

one or more computers and one or more storage devices storing instructions that are operable, when executed by the one or more computers, to cause the one or more computers to perform operations comprising:

obtaining, by one or more computers, a first seed of nucleotides from a reference sequence, wherein the first seed has a length of K nucleotides;

determining, by the one or more computers, that the first seed matches more than a predetermined number of reference sequence locations;

based on determining that the first seed matches more than a predetermined number of reference sequence locations, generating, by the one or more computers, a seed extension tree having a plurality of nodes, wherein each node of the plurality of nodes corresponds to (i) an extended seed that is an extension of the first seed and has a nucleotide length of K^* , wherein K^* is one or more nucleotides greater than K , and (ii) one or more locations, in a seed extension table, that include data describing reference sequence locations that match the extended seed; and

for each node of the plurality of nodes:

storing, by the one or more computers, interval information at a location of the hash table that corresponds to an index key for the extended seed, wherein the interval information references one or more locations in the seed extension table that include data describing reference sequence locations that match the extended seed associated with the node.

Claim 19. The system of claim 18, wherein each of the matching reference sequence locations includes the K nucleotides of the first seed.

Claim 20. The system of claim 18, the operations further comprising:

- obtaining, by the one or more computers, a second seed of nucleotides from the reference sequence that is different than the first seed;
- determining, by the one or more computers, that the second seed does not match more than the predetermined number of reference sequence locations; and
- based on determining, by the one or more computers, that the second seed does not match more than the predetermined number of reference sequence locations:
 - obtaining, by the one or more computers, data describing each of the reference sequence locations that match the second seed; and
 - storing, by the one or more computers, the data describing the reference sequence locations that match the second seed at a second location of the hash table that corresponds to an index key for the second seed.

Claim 21. The system of any one of claims 18 to 20, wherein the one or more locations in the seed extension table that include data describing reference sequence locations that match the extended seed comprises a contiguous interval of locations in the seed extension table that include data describing reference sequence locations that match the extended seed.

Claim 22. The system of any one of claims 18 to 20, wherein the one or more locations in the seed extension table that include data describing reference sequence locations that match the extended seed comprises a non-contiguous interval in an extension table of reference sequence locations that match the extended seed.

Claim 23. The system of any one of claims 18 to 22, wherein obtaining, by a one or more computers, a first seed of nucleotides from a reference sequence, the first seed representing a sequence of nucleotides having a nucleotide length of K nucleotides comprises:

- determining, by the one or more computers, a location of a seed access window within a reference sequence; and

obtaining, by the one or more computers, a subset of the reference sequence that is identified by the seed access window.

Claim 24. The system of claim 23, the operations further comprising:

adjusting, by the one or more computers, the seed extension window K nucleotides forward along the reference sequence to identify a second seed of nucleotides from the reference sequence having a nucleotide length of K nucleotides;

obtaining, by the one or more computers, the second seed from the reference sequence;

determining, by the one or more computers, that the second seed matches more than a predetermined number of reference sequence locations;

based on determining that the second seed matches more than a predetermined number of reference sequence locations, generating, by the one or more computers, a second seed extension tree having a plurality of second nodes, wherein each second node of the plurality of second nodes corresponds to (i) a second extended seed that is an extension of the second seed and has a nucleotide length of K^* , wherein K^* is one or more nucleotides greater than K, and (ii) a second one or more locations, in a second seed extension table, that includes data describing reference sequence locations that match the second extended seed; and

for each second node of the plurality of second nodes:

storing, by the one or more computers, second interval information at a location of the hash table that corresponds to an index key for the second extended seed, wherein the second interval information references one or more locations in the second seed extension table that include data describing reference sequence locations that match the second extended seed associated with the second node.

Claim 25. The system of any one of claims 18 to 24, the operations further comprising:

for each node of the plurality of nodes:

determining, by the one or more computers, whether the node of the seed extension tree is a leaf node; and

based on determining, by the one or more computers, that the node of the extension tree is not a leaf node, storing, by the one or more computers, an extend record in the location of the hash table that corresponds to the index key for the extended seed.

Claim 26. The system of claim 25, wherein the extend record includes one or more instructions that, when executed by the one or more computers, cause the one or more computers to add one or more additional nucleotides to a seed associated with the extend record.

Claim 27. The system of any one of claims 18 to 25, the operations further comprising:
based on determining, by the one or more computers, that the node extension tree is a leaf node, determining, by the one or more computers, to not store an extend record in the location of the hash table that corresponds to the index key for the extended seed.

Claim 28. The system of any one of claims 18 to 25, the operations further comprising:
generating, by the one or more computers, the seed extension table, comprising:
identifying, by the one or more computers, each seed of the reference sequence that matches the first seed; and
storing, by the one or more computers, data identifying the identified seeds in the seed extension table.

Claim 29. The system of any one of claims 18 to 28, the operations further comprising:
sorting, by the one or more computers, the identified seeds in the seed extension table.

Claim 30. The system of any one of claims 18 to 28, the operations further comprising:
generating, by the one or more computers, a hash table installation package, the hash table installation package including instructions that, when processed by one or more computers that receive the hash table installation package, cause the one or more computers to install the hash table in a memory that is accessible by a programmable logic circuit.

Claim 31. The system of claim 30,
wherein the hash table installation package includes the seed extension table, and
wherein the hash table installation package includes instructions instruct (i) the programmable logic circuit or (ii) another computer to store the seed extension table in a memory device that is accessible to the programmable logic circuit.

Claim 32. The system of claim 30 or 31, the operations further comprising:
providing, by the one or more computers, the hash table installation package to another computer.

Claim 33. The system of claim 32, wherein the other computer includes (i) a computer that is configured to communicate with the programmable logic circuit or (ii) the programmable logic circuit.

Claim 34. The system of any one of claims 18 to 32, wherein the one or more computers comprises a plurality of computers.

Claim 35. A non-transitory computer-readable medium storing software comprising instructions executable by one or more computers which, upon such execution, cause the one or more computers to perform operations comprising:

obtaining, by one or more computers, a first seed of nucleotides from a reference sequence, wherein the first seed has a length of K nucleotides;

determining, by the one or more computers, that the first seed matches more than a predetermined number of reference sequence locations;

based on determining that the first seed matches more than a predetermined number of reference sequence locations, generating, by the one or more computers, a seed extension tree having a plurality of nodes, wherein each node of the plurality of nodes corresponds to (i) an extended seed that is an extension of the first seed and has a nucleotide length of K^* , wherein K^* is one or more nucleotides greater than K , and (ii) one or more locations, in a seed extension table, that include data describing reference sequence locations that match the extended seed; and

for each node of the plurality of nodes:

storing, by the one or more computers, interval information at a location of the hash table that corresponds to an index key for the extended seed, wherein the interval information references one or more locations in the seed extension table that include data describing reference sequence locations that match the extended seed associated with the node.

Claim 36. The computer-readable medium of claim 35, wherein each of the matching reference sequence locations includes the K nucleotides of the first seed.

Claim 37. The computer-readable medium of claim 35, the operations further comprising:
obtaining, by the one or more computers, a second seed of nucleotides from the reference sequence that is different than the first seed;
determining, by the one or more computers, that the second seed does not match more than the predetermined number of reference sequence locations; and
based on determining, by the one or more computers, that the second seed does not match more than the predetermined number of reference sequence locations:
obtaining, by the one or more computers, data describing each of the reference sequence locations that match the second seed; and
storing, by the one or more computers, the data describing the reference sequence locations that match the second seed at a second location of the hash table that corresponds to an index key for the second seed.

Claim 38. The computer-readable medium of any one of claims 35 to 37, wherein the one or more locations in the seed extension table that include data describing reference sequence locations that match the extended seed comprises a contiguous interval of locations in the seed extension table that include data describing reference sequence locations that match the extended seed.

Claim 39. The computer-readable medium of any one of claims 35 to 37, wherein the one or more locations in the seed extension table that include data describing reference sequence locations that match the extended seed comprises a non-contiguous interval in an extension table of reference sequence locations that match the extended seed.

Claim 40. The computer-readable medium of any one of claims 35 to 39, wherein obtaining, by a one or more computers, a first seed of nucleotides from a reference sequence, the first seed representing a sequence of nucleotides having a nucleotide length of K nucleotides comprises:

determining, by the one or more computers, a location of a seed access window within a reference sequence; and

obtaining, by the one or more computers, a subset of the reference sequence that is identified by the seed access window.

Claim 41. The computer-readable medium of claim 40, the operations further comprising:

adjusting, by the one or more computers, the seed extension window K nucleotides forward along the reference sequence to identify a second seed of nucleotides from the reference sequence having a nucleotide length of K nucleotides;

obtaining, by the one or more computers, the second seed from the reference sequence;

determining, by the one or more computers, that the second seed matches more than a predetermined number of reference sequence locations;

based on determining that the second seed matches more than a predetermined number of reference sequence locations, generating, by the one or more computers, a second seed extension tree having a plurality of second nodes, wherein each second node of the plurality of second nodes corresponds to (i) a second extended seed that is an extension of the second seed and has a nucleotide length of K^* , wherein K^* is one or more nucleotides greater than K, and (ii) a second one or more locations, in a second seed extension table, that includes data describing reference sequence locations that match the second extended seed; and

for each second node of the plurality of second nodes:

storing, by the one or more computers, second interval information at a location of the hash table that corresponds to an index key for the second extended seed, wherein the second interval information references one or more locations in the second seed extension table that include data describing reference sequence locations that match the second extended seed associated with the second node.

Claim 42. The computer-readable medium of any one of claims 35 to 41, the operations further comprising:

for each node of the plurality of nodes:

determining, by the one or more computers, whether the node of the seed extension tree is a leaf node; and

based on determining, by the one or more computers, that the node of the extension tree is not a leaf node, storing, by the one or more computers, an extend record in the location of the hash table that corresponds to the index key for the extended seed.

Claim 43. The computer-readable medium of claim 42, wherein the extend record includes one or more instructions that, when executed by the one or more computers, cause the one or more computers to add one or more additional nucleotides to a seed associated with the extend record.

Claim 44. The computer-readable medium of any one of claims 35 to 42, the operations further comprising:

based on determining, by the one or more computers, that the node extension tree is a leaf node, determining, by the one or more computers, to not store an extend record in the location of the hash table that corresponds to the index key for the extended seed.

Claim 45. The computer-readable medium of any one of claims 35 to 42, the operations further comprising:

generating, by the one or more computers, the seed extension table, comprising:

identifying, by the one or more computers, each seed of the reference sequence that matches the first seed; and

storing, by the one or more computers, data identifying the identified seeds in the seed extension table.

Claim 46. The computer-readable medium of any one of claims 35 to 45, the operations further comprising:

sorting, by the one or more computers, the identified seeds in the seed extension table.

Claim 47. The computer-readable medium of any one of claims 35 to 45, the operations further comprising:

generating, by the one or more computers, a hash table installation package, the hash table installation package including instructions that, when processed by one or more computers

that receive the hash table installation package, cause the one or more computers to install the hash table in a memory that is accessible by a programmable logic circuit.

Claim 48. The computer-readable medium of claim 47,
wherein the hash table installation package includes the seed extension table, and
wherein the hash table installation package includes instructions instruct (i) the
programmable logic circuit or (ii) another computer to store the seed extension table in a memory
device that is accessible to the programmable logic circuit.

Claim 49. The computer-readable medium of claim 47 or 48, the operations further comprising:
providing, by the one or more computers, the hash table installation package to another
computer.

Claim 50. The computer-readable medium of claim 49, wherein the other computer includes (i)
a computer that is configured to communicate with the programmable logic circuit or (ii) the
programmable logic circuit.

Claim 51. The computer-readable medium of any one of claim 35 to 50, wherein the one or
more computers comprises a plurality of computers.

Claim 52. A method for using a hash table to improve mapping of sample reads to a reference
sequence, the method comprising:

executing, by a mapping and aligning unit, a query of a hash table, the query including a first
seed, wherein the first seed includes a subset of nucleotides that were obtained from a particular read
of the sample reads;

obtaining, by the mapping and aligning unit, a response to the executed query that includes
information stored by a location of the hash table that is determined to be responsive to the query;

determining, by the mapping and aligning unit, whether the response to the executed query
includes (i) an extend record and (ii) an interval record, wherein the interval record identifies a
contiguous set of reference sequence locations, stored in an extension table, that match the first seed;

based on determining, by the mapping and aligning unit, that the response to the executed
query includes (i) an extend record and (ii) an interval record:

determining, by the mapping and aligning unit, whether the extension table is to be accessed to obtain one or more matching reference sequence locations in the extension table that are referenced by the interval record;

based on determining that the extension table is not to be accessed:

determining, by the mapping and aligning unit, whether to store the first information describing the interval record in a memory device as information describing a candidate best interval;

generating, by the mapping and aligning unit, a first extended seed that is an extension of the first seed using the extend record;

generating, by the mapping and aligning unit, a subsequent hash query that includes the first extended seed; and

executing, by the mapping and aligning unit, the subsequent hash query of the hash table.

Claim 53. The method of claim 52, the method further comprising:

based on determining that the extension table is to be accessed:

accessing, by the mapping and aligning unit, the extension table to obtain the one or more matching reference sequence locations in the extension table that are referenced by the interval record; and

adding, by the mapping and aligning unit, the one or more matching reference sequence locations to a seed match set.

Claim 54. The method of claim 52 or 53, the method further comprising:

determining, by the mapping and aligning unit, that the response to the executed query includes one or more matching reference sequence locations; and

based on determining, by the mapping and aligning unit, that the response to the executed query includes one or more matching reference sequence locations:

adding, by the mapping and aligning unit, the one or more matching reference sequence locations to a seed match set.

Claim 55. The method of any one of claims 52 to 54, wherein determining, by the mapping and aligning unit, whether to store the first information describing the interval record in a memory device as information describing a candidate best interval comprises:

determining, by the mapping and aligning unit, that there is not prior information describing an interval record as a candidate best interval for the particular read; and

storing, by the mapping and aligning unit, the first information describing the interval record in the memory device as information describing a candidate best interval.

Claim 56. The method of any one of claims 52 to 55, the method further comprising:

obtaining, by the mapping and aligning unit, a response to the subsequent executed query that includes information stored by a location of the hash table that is determined to be responsive to the query;

determining, by the mapping and aligning unit, whether the response to the subsequent executed query includes (i) a second extend record (ii) a second interval record, or (iii) one or more matching reference sequence locations;

based on determining, by the mapping and aligning unit, that the response to the subsequent executed query includes (i) the second extend record and (ii) the second interval record:

determining, by the mapping and aligning unit, whether the extension table is to be accessed to obtain one or more matching reference sequence locations in the extension table that are referenced by the second interval record;

based on determining that the extension table is not to be accessed:

determining, by the mapping and aligning unit and using one or more heuristic rules, whether second information describing the second interval record or the first information describing the candidate best interval is to be used as the candidate best interval;

generating, by the mapping and aligning unit, a second extended seed that is an extension of the first extended seed using the second extend record;

generating, by the mapping and aligning unit, a third hash query that includes the second extended seed; and

executing, by the mapping and aligning unit, the third query of the hash table that includes the second extended seed.

Claim 57. The method of claim 56, wherein determining, by the mapping and aligning unit and using one or more heuristic rules, whether the second information describing the second interval record or the first information describing the candidate best interval is to be used as the best interval comprises:

selecting either the second information describing the second interval record or the first information describing the candidate best interval record based on a plurality of factors that include (i) a number of matching reference sequence locations returned by each of the interval record and the second interval record, (ii) a predetermined threshold level of reference sequence locations, or (iii) each seed length of the respective seeds that reached the hash locations storing the interval record and the second interval record.

Claim 58. The method of any one of claims 52 to 57, wherein the interval record references one or more locations, in the seed extension table, that include data describing reference sequence locations that match the first seed of the query.

Claim 59. The method of claim 58, wherein the one or more locations, in the seed extension table, that include data describing reference sequence locations that match the first seed of the query comprises:

a contiguous interval, in an extension table, of reference sequence locations that match the first seed of the query.

Claim 60. A system for using a hash table to improve the mapping of sample reads to a reference sequence comprising:

one or more computers and one or more storage devices storing instructions that are operable, when executed by the one or more computers, to cause the one or more computers to perform operations comprising:

executing, by a mapping and aligning unit, a query of a hash table, the query including a first seed, wherein the first seed includes a subset of nucleotides that were obtained from a particular read of the sample reads;

obtaining, by the mapping and aligning unit, a response to the executed query that includes information stored by a location of the hash table that is determined to be responsive to the query;

determining, by the mapping and aligning unit, whether the response to the executed query includes (i) an extend record and (ii) an interval record, wherein the interval record identifies a contiguous set of reference sequence locations, stored in an extension table, that match the first seed;

based on determining, by the mapping and aligning unit, that the response to the executed query includes (i) an extend record and (ii) an interval record:

determining, by the mapping and aligning unit, whether the extension table is to be accessed to obtain one or more matching reference sequence locations in the extension table that are referenced by the interval record;

based on determining that the extension table is not to be accessed:

determining, by the mapping and aligning unit, whether to store the first information describing the interval record in a memory device as information describing a candidate best interval;

generating, by the mapping and aligning unit, a first extended seed that is an extension of the first seed using the extend record;

generating, by the mapping and aligning unit, a subsequent hash query that includes the first extended seed; and

executing, by the mapping and aligning unit, the subsequent hash query of the hash table.

Claim 61. The system of claim 60, the operations further comprising:
based on determining that the extension table is to be accessed:

accessing, by the mapping and aligning unit, the extension table to obtain the one or more matching reference sequence locations in the extension table that are referenced by the interval record; and

adding, by the mapping and aligning unit, the one or more matching reference sequence locations to a seed match set.

Claim 62. The system of claim 60 or 61, the operations further comprising:
determining, by the mapping and aligning unit, that the response to the executed query includes one or more matching reference sequence locations; and
based on determining, by the mapping and aligning unit, that the response to the executed query includes one or more matching reference sequence locations:
adding, by the mapping and aligning unit, the one or more matching reference sequence locations to a seed match set.

Claim 63. The system of any one of claims 60 to 62, wherein determining, by the mapping and aligning unit, whether to store the first information describing the interval record in a memory device as information describing a candidate best interval comprises:
determining, by the mapping and aligning unit, that there is not prior information describing an interval record as a candidate best interval for the particular read; and
storing, by the mapping and aligning unit, the first information describing the interval record in the memory device as information describing a candidate best interval.

Claim 64. The system of any one of claims 60 to 63, the operations further comprising:
obtaining, by the mapping and aligning unit, a response to the subsequent executed query that includes information stored by a location of the hash table that is determined to be responsive to the query;
determining, by the mapping and aligning unit, whether the response to the subsequent executed query includes (i) a second extend record (ii) a second interval record, or (iii) one or more matching reference sequence locations;
based on determining, by the mapping and aligning unit, that the response to the subsequent executed query includes (i) the second extend record and (ii) the second interval record:

determining, by the mapping and aligning unit, whether the extension table is to be accessed to obtain one or more matching reference sequence locations in the extension table that are referenced by the second interval record;

based on determining that the extension table is not to be accessed:

determining, by the mapping and aligning unit and using one or more heuristic rules, whether second information describing the second interval record or the first information describing the candidate best interval is to be used as the candidate best interval;

generating, by the mapping and aligning unit, a second extended seed that is an extension of the first extended seed using the second extend record;

generating, by the mapping and aligning unit, a third hash query that includes the second extended seed; and

executing, by the mapping and aligning unit, the third query of the hash table that includes the second extended seed.

Claim 65. The system of claim 64, wherein determining, by the mapping and aligning unit and using one or more heuristic rules, whether the second information describing the second interval record or the first information describing the candidate best interval is to be used as the best interval comprises:

selecting either the second information describing the second interval record or the first information describing the candidate best interval record based on a plurality of factors that include (i) a number of matching reference sequence locations returned by each of the interval record and the second interval record, (ii) a predetermined threshold level of reference sequence locations, or (iii) each seed length of the respective seeds that reached the hash locations storing the interval record and the second interval record.

Claim 66. The system of any one of claims 60 to 65, wherein the interval record references one or more locations, in the seed extension table, that include data describing reference sequence locations that match the first seed of the query.

Claim 67. The system of claim 66, wherein the one or more locations, in the seed extension table, that include data describing reference sequence locations that match the first seed of the query comprises:

- a contiguous interval, in an extension table, of reference sequence locations that match the first seed of the query.

Claim 68. A non-transitory computer-readable medium storing software comprising instructions executable by one or more computers which, upon such execution, cause the one or more computers to perform operations comprising:

- executing, by a mapping and aligning unit, a query of a hash table, the query including a first seed, wherein the first seed includes a subset of nucleotides that were obtained from a particular read of the sample reads;

- obtaining, by the mapping and aligning unit, a response to the executed query that includes information stored by a location of the hash table that is determined to be responsive to the query;

- determining, by the mapping and aligning unit, whether the response to the executed query includes (i) an extend record and (ii) an interval record, wherein the interval record identifies a contiguous set of reference sequence locations, stored in an extension table, that match the first seed;

- based on determining, by the mapping and aligning unit, that the response to the executed query includes (i) an extend record and (ii) an interval record:

 - determining, by the mapping and aligning unit, whether the extension table is to be accessed to obtain one or more matching reference sequence locations in the extension table that are referenced by the interval record;

 - based on determining that the extension table is not to be accessed:

 - determining, by the mapping and aligning unit, whether to store the first information describing the interval record in a memory device as information describing a candidate best interval;

 - generating, by the mapping and aligning unit, a first extended seed that is an extension of the first seed using the extend record;

generating, by the mapping and aligning unit, a subsequent hash query that includes the first extended seed; and

executing, by the mapping and aligning unit, the subsequent hash query of the hash table.

Claim 69. The computer-readable medium of claim 68, the operations further comprising: based on determining that the extension table is to be accessed:

accessing, by the mapping and aligning unit, the extension table to obtain the one or more matching reference sequence locations in the extension table that are referenced by the interval record; and

adding, by the mapping and aligning unit, the one or more matching reference sequence locations to a seed match set.

Claim 70. The computer-readable medium of claim 68 or 69, the operations further comprising:

determining, by the mapping and aligning unit, that the response to the executed query includes one or more matching reference sequence locations; and

based on determining, by the mapping and aligning unit, that the response to the executed query includes one or more matching reference sequence locations:

adding, by the mapping and aligning unit, the one or more matching reference sequence locations to a seed match set.

Claim 71. The computer-readable medium of one of claims 68 to 70, wherein determining, by the mapping and aligning unit, whether to store the first information describing the interval record in a memory device as information describing a candidate best interval comprises:

determining, by the mapping and aligning unit, that there is not prior information describing an interval record as a candidate best interval for the particular read; and

storing, by the mapping and aligning unit, the first information describing the interval record in the memory device as information describing a candidate best interval.

Claim 72. The computer-readable medium of one of claims 68 to 71, the operations further comprising:

- obtaining, by the mapping and aligning unit, a response to the subsequent executed query that includes information stored by a location of the hash table that is determined to be responsive to the query;

- determining, by the mapping and aligning unit, whether the response to the subsequent executed query includes (i) a second extend record (ii) a second interval record, or (iii) one or more matching reference sequence locations;

- based on determining, by the mapping and aligning unit, that the response to the subsequent executed query includes (i) the second extend record and (ii) the second interval record:

 - determining, by the mapping and aligning unit, whether the extension table is to be accessed to obtain one or more matching reference sequence locations in the extension table that are referenced by the second interval record;

 - based on determining that the extension table is not to be accessed:

 - determining, by the mapping and aligning unit and using one or more heuristic rules, whether second information describing the second interval record or the first information describing the candidate best interval is to be used as the candidate best interval;

 - generating, by the mapping and aligning unit, a second extended seed that is an extension of the first extended seed using the second extend record;

 - generating, by the mapping and aligning unit, a third hash query that includes the second extended seed; and

 - executing, by the mapping and aligning unit, the third query of the hash table that includes the second extended seed.

Claim 73. The computer-readable medium of claim 72, wherein determining, by the mapping and aligning unit and using one or more heuristic rules, whether the second information describing the second interval record or the first information describing the candidate best interval is to be used as the best interval comprises:

selecting either the second information describing the second interval record or the first information describing the candidate best interval record based on a plurality of factors that include (i) a number of matching reference sequence locations returned by each of the interval record and the second interval record, (ii) a predetermined threshold level of reference sequence locations, or (iii) each seed length of the respective seeds that reached the hash locations storing the interval record and the second interval record.

Claim 74. The computer-readable medium of one of claim 68 to 73, wherein the interval record references one or more locations, in the seed extension table, that include data describing reference sequence locations that match the first seed of the query.

Claim 75. The computer-readable medium of claim 74, wherein the one or more locations, in the seed extension table, that include data describing reference sequence locations that match the first seed of the query comprises:

a contiguous interval, in an extension table, of reference sequence locations that match the first seed of the query.